Supplementary Materials for Unsupervised Domain Adaptation with Hierarchical Gradient Synchronization

Lanqing Hu[1,2] Meina Kan[1,2] Shiguang Shan[1,2,3] Xilin Chen[1,2]

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, 200031, China

lanqing.hu@vipl.ict.ac.cn {kanmeina,sgshan,xlchen}@ict.ac.cn

1. Evaluation on Partial Domain Adaptation

Our method is also applicable for partial domain adaptation as it ensures a favorable alignment of each local class. So, we further compare our method with those partial domain adaptation approaches under Office-31-PDA for more comprehensive investigation. The evaluation results are shown in Table 1. Specially, for reducing the class unbalance problem in partial domain adaptation, we follow PADA [2] to down weight those uncommon classes learned from the model during training process.

DAN [7], DANN [4], ADDA [12], RTN [9] and LEL [10] are methods for standard domain adaptation and they are straightforwardly applied to the partial domain adaptation for comparison. These methods all align domain distribution in a global manner and without handling the negative transfer in partial domain adaptation. Hence, the performance of them are not so good, especially DAN [7], DANN [4] and ADDA [12] even perform worse than the source model without any adaptation. MCDDA [11], DupGAN [6] and CDAN [8] considers category discrimination during adaptation leading to better domain adaptation performance. As can be seen from Table 1, our method GSDA obtains the state-of-the-art performance, and even outperforms SAN [1], PADA [2] and ETN [3] that are particularly designed for partial domain adaptation. This superior improvement can be safely attributed to the consideration of the consistency between global and local distribution alignment, which can naturally only calibrate those common categories and ignore those categories that only appear in source domain.

In experiment of partial domain adaptation, we also find that our method is more stable and robust than the previous related works. Figure 1 displays the testing accuracy of different methods w.r.t. different training iterations in adaption from A31 to W10. As can be seen, the testing accuracy of DANN, which is a global alignment method, decreases dramatically during training because only aligning the global distribution will result in negative transfer on the uncommon classes. The accuracies of SAN with class-wise alignment and SAN+PADA with both global and class-wise alignment are more stable but still have slight drops. Surprisingly, the testing accuracy of our method grows up steadily in the whole training process demonstrating its robustness and effectiveness.

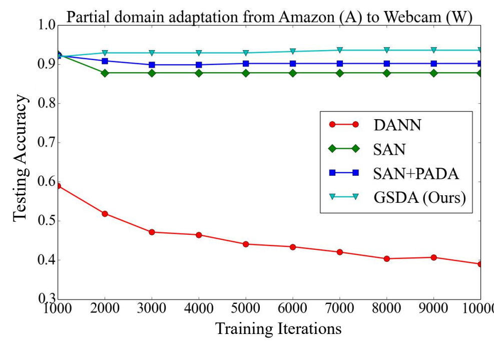

Figure 1. The comparison of testing accuracy of the existing methods and our GSDA w.r.t. models from different training iterations on experiment of A31 $\rightarrow$ W10. Best viewed in color.

2. Visualization of the Distribution Calibration Process of GSDA

In this part, the detailed optimizing process on selected three classes from source and target domains is visualized for better understanding of the proposed method. Figure 2 shows the aligning process of two methods, our *GSDA*, and a baseline *GSDA w/o gs* (i.e., GSDA without hierarchical gradient synchronization). As can be seen, initially both methods misalign the category of ring binder marked in diamonds with the category of notebook in target domain. Unfortunately, the *GSDA w/o gs* makes the optimization go further along this wrong direction as only this way can make its objective loss continually becomes smaller, and

Table 1. Object classification accuracy of partial domain adaptation tasks on Office-31-PDA (ResNet50). All methods follow the same settings, so most results are directly from the original works except those ones marked with stars which are tuned using the released codes.

Method	A31 → W10	D31 → W10	W31 → D10	A31 → D10	D31 → A10	W31 → A10	Avg
ResNet50 [5]	54.5	94.6	94.3	65.6	73.2	71.7	75.6
DAN [7]	46.4	53.6	58.6	42.7	65.7	65.3	55.4
DANN [4]	41.4	46.8	38.9	41.4	41.3	44.7	42.4
ADDA [12]	43.7	46.5	40.1	43.7	42.8	46.0	43.8
RTN [9]	75.3	97.1	98.3	66.9	85.6	85.7	84.8
LEL [10]	73.2	93.9	96.8	76.4	83.6	84.8	84.8
MCDDA* [11]	67.8	96.6	99.4	70.1	70.4	73.8	79.7
DUPGAN* [6]	92.5	96.3	**100.0**	89.2	91.2	92.3	93.6
CDAN* [8]	83.4	98.3	99.4	89.2	93.8	92.3	92.7
PADA [2]	86.5	99.3	**100.0**	82.2	92.7	95.4	92.7
SAN* [1]	92.2	**100.0**	98.7	93.0	95.3	94.8	95.7
ETN [3]	94.5	**100.0**	**100.0**	95.0	**96.2**	94.6	96.7
GSDA (Ours)	**97.6**	**100.0**	**100.0**	**97.5**	96.0	**95.7**	**97.8**

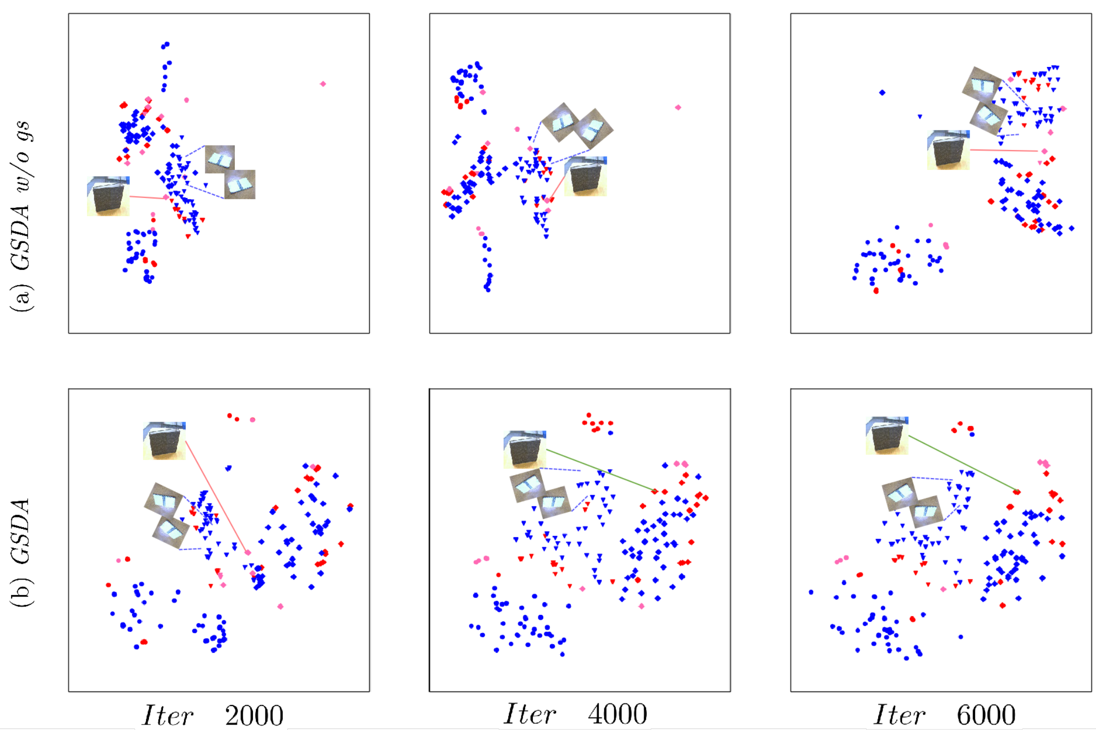

Figure 2. The feature (f) distribution of several categories during training process in Office-31-DA $A \rightarrow W$ adaptation: (a) distribution of GSDA w/o gs with some misalignment, and (b) distribution of *GSDA* with better alignment. The blue points denote samples from source domain. The red and pink points denote samples from target domain, that are aligned correctly and incorrectly respectively. The three columns represent distribution of the selected three categories in iteration 2000, 4000 and 6000, respectively. The same category from two domains are marked with the same shape and different categories are marked with different shapes. Best viewed in color.

not surprisingly the two categories become even closer in the subsequent iterations as shown in Figure 2(a). This misalignment will cause inconsistency between local discriminators and global domain discriminator, which pushes our GSDA method to penalize this wrong optimization and pull the boundary sample back to the right path, ensuring correct category alignment as shown in Figure 2(b).

3. Sensitivity of on Hyper-Parameter Selection

We investigate the sensitivity of the hyper-parameters of our method, including $\alpha = \{0.02, 0.05, 0.2, 0.5\}$ for balancing the entropy loss, $\beta = \{0.5, 1.0, 5.0, 10.0\}$ for balancing the gradient synchronization, and the number of groups $b = \{3, 4, 5, 6, 13\}$. The performance of our method with these different hyper-parameters changes less than 1.0%, demonstrating that our method is not so sensitive to the hyper-parameters. So we present the best performance shown in Tables 2,3,4 with the corresponding hyper-parameters introduced in Section 3.1 in the manuscript.

References

[1] Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Michael I. Jordan. Partial transfer learning with selective adversarial networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2

[2] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *European Conference on Computer Vision (ECCV)*, 2018. 1, 2

[3] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2

[4] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, and et al. Domain-adversarial training of neural networks. *Journal of Machine Learning Research (JMLR)*, 17(1):2096–2030, 2016. 1, 2

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[6] Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. Duplex generative adversarial network for unsupervised domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2

[7] M. Long, Y. Cao, J. Wang, and M. I. Jordan. Learning transferable features with deep adaptation networks. In *International Conference on Machine learning (ICML)*, 2015. 1, 2

[8] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *Neural Information Processing Systems (NeurIPS)*, 2018. 1, 2

[9] M. Long, H. Zhu, J. Wang, and M. I. Jordan. Unsupervised domain adaptation with residual transfer networks. In *Neural Information Processing Systems (NeurIPS)*, 2016. 1, 2

[10] Zelun Luo, Yuliang Zou, Judy Hoffman, and Li F Fei-Fei. Label efficient learning of transferable representations acrosss domains and tasks. In *Neural Information Processing Systems (NeurIPS)*, 2017. 1, 2

[11] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2

[12] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 2